A Pyramid of Decision Approaches

Paul J.H. Schoemaker   **J. Edward Russo**

Nothing is more difficult, and therefore more precious, than to be able to decide.
—Napoleon Bonaparte (*Maxims*, 1804)

Most managers still make decisions based on intuition, despite the risks. It's true that computers have improved information gathering and display and that some routine decisions, such as credit applications and inventory ordering, can be automated. But most managerial decisions are still disturbingly immune to technological and conceptual advances.

Managers know that decision making is more critical than ever; with global competition, managers are competing against the best of the best. Recent decision research has offered insights into improving managerial decisions that were not available even a decade ago. But how can you incorporate some of those insights into the decisions that you, your colleagues, and your subordinates make?

There are four general approaches to decision making, ranging from intuitive to highly analytical.

Intuition

> Many complain about their memory, few about their judgment.
> —La Rochefoucauld

Intuition is quick and easy. It's hard to dispute decisions based on intuition because the decision makers can't articulate the underlying reasoning. People just know they're right, or they have a strong feeling about it, or they're relying on "gut feel." Of course, if such a decision turns out to be wrong, the decision maker has no defense.

Intuition can sometimes be brilliant. When based on extensive learning from past experience, it may truly reflect "automated expertise."[1] Some managers are so familiar with certain situations that they grasp the key issues instantly and nearly automatically. However, they may have great difficulty explaining their intuition. How much credibility can we give such decisions? Decision research has revealed two common flaws in intuitive decision making: random inconsistency and systematic distortion.

Inconsistency—Nine radiologists were independently shown information from 96 cases of suspected stomach ulcers and asked to evaluate each case in terms of the likelihood of a malignancy.[2] A week later, after these X-ray specialists had forgotten the details of the 96 cases, they were presented with the same ones again, but in a different order. A comparison of the two sets of diagnoses showed a 23% chance that an opinion would be changed.

People often apply criteria inconsistently. They don't realize how much memory failings, mental limits, distractions, and fatigue can influence their judgments from one time to the next. Not one of the radiologists was perfectly consistent, and the worst ones were inconsistent to an alarming extent. Note that these were highly trained professionals making judgments central to their work. In addition, they knew that their medical judgments were being examined by researchers, so they probably tried as hard as they could. Still, they made significant errors.

One reason people make different decisions on different days is that they don't test themselves for inconsistency. They believe they're consistent and don't build in safeguards. Indeed, few experiments like the one with the radiologists have been conducted. We know of none involving managers.

We asked 128 MBA students to predict the grade point average (GPA) of 50 past students. These were listed in random order (without names) and described only by the standard information in their completed applications, such as test scores, college grades, and so on. Three weeks later, we asked the 128 students to repeat this task and challenged them to be as consistent with their initial predictions as possible. They performed slightly worse

than the radiologists, *even though they knew they were being tested on consistency.* Imagine the level of error that creeps in when we're not watching for it.

Distrust intuition. Random inconsistency isn't just an isolated danger for certain experts operating in especially difficult situations. It is a widespread shortcoming in most people and in most work situations. Inconsistency is a constant and hidden threat to good decision making.

Distortion—People often systematically under- or overemphasize certain pieces of information. We tend to overemphasize the most recent information we have received. That's why the last person to get the boss's ear has the most influence and why the closing arguments of a trial can sway the jurors. Sometimes we respond to the first information we receive. Sales people know this, and they try to beat competitors to new customers. Anyone going to a job interview knows this and tries to make a good first impression. We also tend to pay more attention to information that is readily available. People tend to be more afraid of highly reported accidents such as airplane crashes, earthquakes, and nuclear meltdowns than the more common but underreported ones such as at-home accidents, drowning, and electrocution. Furthermore, each of these judgmental distortions is amplified when people place, as they typically do, too much confidence in their intuitive judgment.[3]

Even when inconsistency is eliminated, distortion leads to suboptimal judgments. Securities analysts were asked to predict the earnings growth of certain U.S. companies.[4] At the same time, a statistical model was developed to predict earnings growth based solely on past earnings. The analysts had access to the same information the model had, but their predictions had a mean correlation of only .23 with the actual earnings.[5] The computer forecasts scored .59. When inconsistency was removed from the analysts' predictions (with another simple regression model), the mean correlation increased to .29, better but still far short of the statistical model. The gap between .29 and .59 reflects the systematic distortion.

Sometimes intuition is the only option. When time is short or when key aspects of the situation are hard to quantify (e.g., the quality of artistic works or fine wines), more systematic decision methods may not be feasible. But intuition's successes are exaggerated and its risks underappreciated.

Challenge yourself, your colleagues, and your subordinates to articulate the reasoning underlying decisions. Make up a test that's appropriate for your job—reviewing applications, estimating sales, or predicting completion times. Take it twice, with enough time between tests to forget the original answers. You'll be surprised by how inconsistently you apply your own criteria. Then consider the following, more systematic procedures.

Rules

> Rules are for the obedience of fools and the guidance of wise men.
> —David Ogilvy

We often use rules to sort information. Some rules are specific to industries or occupations; others are generic. Decisions based on rules are somewhat more accurate than wholly intuitive ones. Rules are quick and often clever ways to approximate an optimal response without having to incur the cost of a detailed analysis. Like intuition, rules are fast and often easy to apply. Unlike intuition, they can be articulated and applied consciously. However, people don't always use rules judiciously, and we often don't realize their inherent distortions. In that blindness lurks the danger.

Industry- and Occupation-Specific Rules—Thousands of rules determine when we change price, replace parts, launch a new product, sell a property, and even hire people. In Exhibit 1 we list a number of rules actually used by managers. Try to assess each rule's strengths and weaknesses. (We have indicated some of the important limitations.) Often these guidelines are *golden rules*, honed and tested through time to best balance effort and accuracy. At times, however, it pays to review whether they still hold true.

When the environment has changed, due to deregulation, new technologies, shifts in consumer preferences or whatever, it is likely that some of the old rules have become outdated. Trammel Crow made a fortune in commercial real estate in Texas by breaking the sacred rule that warehouse and office space should be built only after tenants have signed up. Building on speculation positioned him well for the boom years. Ironically, when the local economy went bust, rigid abidance by his new rule nearly ruined his successors.

Make a list of the rules-of-thumb in your industry and company and encourage others to do the same. Take one rule and think of a situation in which using that rule would produce a good decision. Think of a situation in which the rule led to a bad decision and explain why. What would be the most disastrous application of the rule? Now improve the rule and test it in a pilot project or simulation. In Exhibit 2, we show how this process applies to two rules regarding auditing and pricing.

Generic Rules—People apply a number of generic rules to decisions. The *dictionary rule* is a common one. Suppose you have to select one of several law firms or advertising agencies. A simple strategy is to rank them one attribute at a time, starting with the most important factor. Many managers consider word-of-mouth recommendation the most important criterion. Start by grouping those firms that have all received strong recommendations through the grapevine. Then interview this top tier. In the second group, place those firms that did well in the interview (i.e., those meeting your

Exhibit 1. Actual Rules Used by Managers

- **Restaurant Pricing:** Mark food up three times direct cost, beer four times, and liquor six times. Direct food cost should be no more than 35% of food sales.
  [*Danger:* Ignores labor cost differences and local competitive conditions.]

- **Computer Sales Prospecting:** Seriously pursue a sales prospect only if the prospect's budget for purchasing the computer has already been approved, our product offers some unique benefit, our firm is viewed as a qualified vendor, and the order will be placed within the next six months.
  [*Danger:* Ignores prospects that fail to meet one criterion, but barely, such as a prospect that plans to place a large order in seven months.]

- **Evaluating Acquisitions:** Purchase if and only if the target's estimated after-tax earnings in year 3 (after the purchase) exceed 12% of the purchase price. (This is the rule of a major U.S. company whose CEO made his reputation on an acquisition that yielded 12% after tax in year 3. He now sees a lot of proposals with earnings just over 12%. What he may see less clearly is how the numbers were cooked to meet his rule.)
  [*Danger:* Insensitive to exact income profile over time; hostile to long-term payoffs.]

- **Pricing Seasonal Clothing:** Mark up the wholesale price by 60% and discount the retail price every two weeks by 20% until the entire inventory is gone.
  [*Danger:* Ignores competitors' prices and the special characteristics of each product class.]

- **Conducting Legal Research:** When an issue needs research, tell a law clerk to spend six hours in the library and then report back.
  [*Danger:* Results in overbilling and adverse reaction from clients with small legal issues.]

- **Washington Hotel Booking:** Seven days prior to date accept up to 50 rooms overbooking (on top of 724 rooms available); one day prior to date, accept up to 20 rooms being oversold (used by a well-known Washington, D.C., hotel).
  [*Danger:* Inappropriate with big convention in town when all hotels are over-booked.]

- **IBM Computer Leasing:** Assume a 12% residual value after five years and a debt rate of prime plus one percent for pricing a five-year lease of a new mainframe computer.
  [*Danger:* Ignores credit worthiness of lessee as well as factors changing residual value.]

- **Law Review Editing:** Look at cases cited in the footnotes. If old, so is the point being made. It is probably moot, so don't waste your time reviewing further.
  [*Danger:* Will miss incisive papers that reinterpret classic cases in a new light.]

- **Setting Production Costs:** Estimate the weight of the plastic part, calculate the cost of the material, and then multiply this dollar figure by two for processing, assembly, shipping, etc.
  [*Danger:* Underestimates cost of unique parts with small converter base.]

Exhibit 1. Actual Rules Used by Managers (*continued*)

- **Selecting Ski Resorts:** Select lodge with highest skier traffic; if tied (i.e., within 10%), pick one with the most cooperative management (used by convention and promotion organizer to select sites most conducive to sales events).

  [*Danger:* Ignores ease of travel access as well as quality of rooms, service, food, etc.]

- **Evaluating Bank Teller Performance:** Must process at least two hundred transactions per day, have fewer than four clerical errors per day, and have fewer than five days per month when the cash balance and cash register contents do not match.

  [*Danger:* Discourages high-quality service for elderly or handicapped persons and for new customers still learning how to bank].

- **Bookstore Ordering:** If author and title are not familiar and book is not slated for big review, order 10 copies. Never let inventory drop below two copies.

  [*Danger:* Ignores seasonality (Christmas) and local demand or interest in topic or author.]

- **Software Programming:** When in doubt throw it out; don't waste time trying to patch up someone else's computer program (used within a data-processing center).

  [*Danger:* Overlooks length and complexity of program as well as the nature of the flaw.]

- **Banquet Staffing:** Staff one server per thirty guests if catering a sitdown banquet function and one per forty guests for a buffet.

  [*Danger:* Ignores that serving lobster is more labor intensive than serving chicken; it also ignores that some conventions run on a much tighter time schedule and can ill afford delay.]

- **Exporting Products:** Ship the steel product as long as the contribution margin is positive (used by a Japanese manufacturer serving both foreign and domestic markets).

  [*Danger:* May ship product overseas when domestic demand, which has higher margins, is at capacity, thereby failing to receive the highest contribution margin attainable.]

needs). If ties still remain after this second screen, you might create a third cut by ranking the firms on their fees or bids.

The dictionary rule is commonly used in business. A company might first rank its projects on the basis of expected returns (in rounded percents) and only in cases of ties or near ties consider other attributes such as risk and strategic fit. We call this strategy the dictionary rule because it ranks items the same way a dictionary does: one criterion (i.e., letter) at a time. This rule obviously gives enormous importance to the first attribute and therefore only makes sense if there is a dominant attribute.

But what if your options are not all available at the same time? Suppose you need to make a yes-no decision as credit applications come in. The

Exhibit 2. A Systematic Approach to Evaluating Your Rules

Here are six steps to help you decide if your favorite rule is ready for an overhaul. Two cases illustrate each step.

6 Steps	Case A. Auditing Interest Income	Case B. Pricing a Walkie-Talkie in New Midwest Markets
1. Identify an important rule-of-thumb or short-cut calculation in your firm	Interest applied to end-of-month asset levels should add up to yearly total.	Price by finding closest match of this product in comparable known markets.
2. Give an example of where this rule comes close to the correct answer.	End-of-month balances fairly reflect average monthly balances over the year.	Criteria used to establish matches reflect key demand factors in the new area.
3. Give an actual example of where it failed badly and explain why.	Deposits are received mostly early in the month or withdrawals occur unevenly.	Indianapolis seems like St. Paul but is not growing as fast, which matters here.
4. Construct cases where the rule would produce disastrous results (to understand its limits).	Clerk intends to defraud and thus makes sure month-end figures match year total.	Competitor knows this pricing rule and focuses on markets where it is too high.
5. Generate possible improvements of the rule (from in-house, competitors).	Take several random dates within each month and use those as averages.	Study markets where pricing was clearly off and then develop better criteria.
6. Test the new rules, either in real-world pilot settings or via simulation.	Compare old and new rule on past fraudulent cases.	Try both pricing schemes in new test markets or in ones with strong competitors.

threshold rule allows you to screen each applicant against preset criteria and approve the loan only if all are met. For example, it might be stipulated that a loan will only be granted if:

- the person has no record of payment defaults *and*
- at least 50% of current income is uncommitted *and*
- the person has lived at least a year at the present address *and*
- the person has been at least a year in the present job *and*
- the person's occupation is at least skilled laborer.

Although such a rule is useful, it is too unforgiving. Someone who passes the criteria except for one payment default will not get a loan, which

will likely be the lender's loss. The key questions for the lender are: how many good applicants are turned down; and how many bad ones are accepted because of this rule.

The threshold rule is often used to decide what sort of house to purchase, what car to buy, or what person to hire. Most companies require that investment projects have projected financial returns that pass a preset "hurdle" rate. That's a threshold rule.

Generic rules are often worthy attempts to gain speed and accuracy. They do eliminate random inconsistency and greatly simplify complex tasks. One study found that the dictionary and threshold rules yield about 80% and 30%, respectively, of the accuracy attained by optimal rules.[6] Of course, accuracy rates vary considerably depending on the specific criteria and cutoffs.

But the problem with rules is that they don't take into account all of the relevant information and they don't allow superior performance on some attributes to make up for poor performance on others. So scrutinize the rules you use for the information they leave out and the attributes they emphasize at the expense of others. If you don't recognize the distortions in your rules, you can bet your competitors will see them (see Exhibit 3). The bottom line is to know when and when not to use rules. If you need to consider a more complete set of factors, try importance weighting.

Importance Weighting

The whole trick is to know what variables to look at and then know how to add.
—R. Dawes and B. Corrigan[7]

As we consider the factors that influence a decision, we typically give some factors more weight than others. Importance weighting techniques allow us to articulate those weights, test them, and use them for future decisions. This way you develop a model for applying your own intuitive criteria more consistently and effectively.

Suppose you are judging MBA applicants. You could simply read the application folders and decide to accept or reject each applicant. But, like the students and the radiologists, you would probably give different answers from one week to the next. You could rank the applicants based on a dictionary or threshold rule, but you don't want to neglect any important information. You decide to use importance weighting.

Your first task is to identify and quantify the factors you will use to make your decision. You develop a list of the relevant factors, such as the quality of the personal essay, the selectivity of the applicant's undergraduate institution, undergraduate major, college GPA, work experience, and scores on the Graduate Management Admission Test (GMAT). Some of these factors are already quantified, such as the GPA and GMAT scores. To quantify the other factors, you rate them on some numerical scale (or have an independent

Exhibit 3. How a Smart Cookie Crumbled

A well-known U.S. food company is mysteriously losing market share in five product categories. The culprit: a dictionary rule. The rule, which is used to make decisions about new product formulations in areas such as cookies, goes like this:

> Replace the current product with a cost-reduced version if:
> 1. Consumers do not rate the cheaper version lower in overall satisfaction and
> 2. The new formulation is cheaper per unit sold.

The company uses sophisticated consumer acceptance tests, focusing on overall taste, texture, visual appeal, and so forth, to compare the current and new versions. It also uses appropriate statistical tests and introduces the new version only if the rating differences are not statistically significant. Nonetheless, market share has declined in five categories since multiple cost-reduced versions have been introduced over a period of several years. The declines are not due to industry trends, competitor action, or shifts in consumer preferences. What is happening?

The answer lies in the danger inherent in the dictionary rule. Consumers didn't notice the loss of quality from one change to the next, but the cumulative effect of *several* changes has made a noticeable difference. The company has tested each new version only against the current one and not against any of the previous formulations. It has failed to detect the gradual decline in quality.

The scientific techniques this company is using, such as statistical tests, are embedded in a decision system that is itself flawed. One brand manager, who discovered and persistently complained about the rule, was ignored. He has resigned in frustration. The company still hasn't changed the rule. It *has* noticed the decline in sales, and it issued a warning to brand managers not to be overly zealous in reducing costs at the expense of product quality and brand image. Yet the rule remains in place today, and the company will continue to lose market share.

expert rate them for you). In Table 1 we list these factors and some appropriate rating scales. Now you rate each applicant on these scales.

The second task is to weigh the importance of these factors relative to one another. For instance, you might decide that the personal essay should count as 5% of the decision, the undergraduate major should count as 10%, and so on, up to 100% (see Table 1, third column).

The third step is to multiply each factor's score by the appropriate weight and add all the weighted scores to come up with an overall score for the applicant. For instance, an excellent personal essay would translate into five points (a score of 100 multiplied by 5%). This step can be done easily in a spreadsheet or even with a calculator.

The value of this technique is obvious. You are forced to identify the factors you are using to make the decision and to articulate which factors are most and least important. Importance weighting techniques make intuitive judgments visible and open to examination, by you and by others. And

Table 1. Evaluating MBA Applicants with Importance Weighting

Factor	Rating Scale						Weight
Personal essay	Poor 0	Weak 25	Average 50	Strong 75	Excellent 100		5%
Selectivity of undergrad. institution	Least 0	Next to least 20	Below Avg. 40	Avg. 60	Above Avg. 80	Highest 100	20%
Undergrad. major	Other 0	Science 50	Business 100				10%
College GPA	2.0 0	3.0 50	4.0 100				25%
Work experience	None 0	Some 25	Medium 50	Much 75	Most 100		10%
GMAT verbal	0 . . . . 0		100 percentile 100				10%
GMAT quant.	0 . . . . 0		100 percentile 100				20%

they offer a complete use of the available information, whereas rules short circuit the process. Note that you can use the same rating scheme and weights next year, when it's time to judge applicants again.

Assigning Weights—The heart of the technique is the assignment of importance weights. There are several ways to do this. The easiest way is to allocate a hundred points across the factors. Consider all the factors and intuitively decide how much weight to give each one. Of course, this method may be inaccurate. A person may assign different weights from one time to the next. One way to guard against such random error is to compare pairs of factors. The quantitative GMAT is how many times more important than the verbal GMAT? Work experience is how many times more important than the personal essay? After judging every possible pair of factors, you will need to use a statistical technique to average out the inconsistencies.[8]

But there are other reasons why managers may not want to use the simple approach. If you ask your employees to assign a hundred points to the factors they consider when making a hiring decision, they may not be willing to reveal their actual preferences. They may assign weights that are politically correct or expedient. Or they may consider it demeaning or meaningless to capture their expertise in a set of simple weights.

Sometimes being overly explicit about one's weighting scheme can be detrimental, as Ford Motor Company discovered in its celebrated Pinto lawsuit. Ford's managers had carefully calculated that the cost of adding a reinforced gas tank would not be justified by the expected number of lives saved from rear-end collision fires. An internal memo rejected a safety improvement that around 1970 cost $11 per car, figuring that the savings of $49.5 million in fewer deaths and injuries was not worth the $137 million it would cost to add this safety feature to 11 million cars and 1.5 million light trucks. In making this tradeoff, Ford valued saving a human life at $200,000 (in 1970 dollars) and avoiding the typical injury at $67,000.[9] Putting a price tag on human life hurt Ford with the jury. However, such judgments are unavoidable: either they are made intuitively or explicitly. And in some cases it may help you in court if, say, your personnel or credit decisions use formulae that explicitly exclude criteria deemed illegal (such as gender, race, or geographic area).

Suppose people don't feel comfortable stating their importance weights outright, either because they don't trust themselves to be accurate or they don't wish to expose their own "importance policy" to the scrutiny of others. How can you nonetheless discover their weights? You might simply request such persons to rate intuitively a number of cases (such as applicants, projects, or budget proposals), using an overall attractiveness scale. In making such judgments, the person *implicitly* assigns more weight to some attributes than others. A technique called regression analysis can infer the weights the decision maker appears to have been using to arrive at his or her ratings.

For instance, a gifted claims handler, with an excellent nose for sniffing out fraudulent cases, was about to retire from her insurance company. She had that rare ability to make good intuitive decisions—decisions based on "automated expertise." Unfortunately, she couldn't spell out how she did it. All she could say was that she looked at such factors as lack of adequate support data, valuable property that did not fit the insured's income level, evasiveness in the police report, financial difficulty such as loss of a job, personal problems like divorce, and frequent or suspicious past claims. By asking the adjuster to rate a wide cross section of applications for fraud potential, the company could statistically infer what weights she used and thereby capture valuable expertise before it left the company.

Note that this is a rather different approach from the previous ones discussed. It determines weights *indirectly*. The decision makers do what they normally do, namely, make judgments about complete cases. A regression program analyzes the judgments and figures out the weights *implicitly* assigned to each component.

How do you decide which method to use—the direct or indirect? Experimental and analytic evidence suggests that for most tasks the simple technique of allocating points directly is sufficient.[10] But some situations lend

themselves to the indirect method: when someone, such as the claims handler, makes accurate judgments but cannot explain how or when biases and prejudices are involved and the decision makers won't reveal them directly. Or perhaps you want to test whether a promising subordinate could take over for a manager. Perform a regression analysis on recent decisions made by the subordinate and the manager, then compare the weights on given attributes. The closer the match, the more likely the subordinate will make decisions that are similar to those the manager makes.

Both direct and indirect weighting procedures are subjective; they are based on intuitive judgments. Can *objective* weights be determined? Yes— when you have good archival data of actual outcomes, and when you are confident that present outcomes are not substantially different from those in the past. Models can be built that measure the relationship of weighted attributes to actual outcomes. For instance, look at past loan applications and create a model that, by weighting the applicant attributes, "predicts" the correct results—repayment or default. Then apply the weights that predicted repayment to present applicants.

The model-building approach to decision making—whether subjective or objective—is a clever but counterintuitive way to improve any expert's judgments. Yet it often succeeds as the earlier example with security analysts showed.[11] The intuitive predictions scored only .23 in terms of their correlation with actual earnings. The subjective model scored .29 (which is significantly better) and the objective .59. But once you have built a model, why use the expert? Why not use the model?

Bootstrapping—This process of determining factors and assigning weights has a value that goes beyond any immediate decision. The combination of factors and subjective weights constitutes a model for the decision making process of a given expert. Once you have created that model, you have the opportunity to replace the expert with the model. And the model will likely outperform the expert.[12]

This increase in performance is called "bootstrapping," for obvious reasons: the model is derived from the expert's own use of the available criteria, but it improves decisions based on those criteria. This happens because the model is not plagued by distractions, fatigue, boredom and all the other factors that make us human. The model applies the expert's insights consistently (without using any additional information).

In Table 2, we list the findings of some major studies on bootstrapping. In these cases, the researchers actually knew the correct value of the variables to be predicted, but the participating professionals did not. For example, the study dealing with the life expectancy of terminally ill cancer patients used cases taken from past records. The researcher knew how long the cancer patients had actually lived. The doctors read disguised cases with the dates of death removed. In all of the cases, the model based on the

Table 2. Some Other Bootstrapping Studies

Type of Prediction Task Studied	Type of Subjects	Intuitive Prediction	Model's Prediction	Bootstrap Effect	No. of Subjects	Judgments Per Subject
Excess Returns of Stocks	Security Analysts	0.01	0.01	0.00	21	18
Applications to Graduate School	Admission Officers	0.19	0.25	0.06	1	111
Life Expectancy of Cancer Patients	Medical Doctors	−0.01	0.13	0.14**	3	186
Earnings Growth of Companies	Security Analysts	0.23	0.29	0.06	5	35
Mental Illness Using MMPI Test	Clinical Psychologists	0.28	0.31	0.03***	29	861
Grades in Psychology Course	Graduate Students	0.48	0.56	0.08***	8	50
IQ Scores	Clinical Psychologists	0.51	0.51	0.00	15	100
Bankruptcy Using Financial Ratios	Loan Officers	0.50	0.53	0.03**	43	70
Student Ratings of College Teacher	Other Students	0.35	0.56	0.21	1	16
Success of Life Insurance Agents	Agency Managers	0.13	0.14	0.01	16	200
IQ Scores	College Students	0.47	0.51	0.04*	10	78
Freshman GPAs	Other Students	0.33	0.50	0.17***	98	90
Graduate Admissions	Other Students	0.37	0.43	0.06#	40	90
Changes in Stock Prices	MBA Students	0.23	0.28	0.05**	47	50
Induced Value of Ellipses	Students	0.84	0.89	0.05	6	180
Mean		**0.33**	**0.39**	**0.07**	**23**	**142**

Note: Asterisks denote statistical significance as follows: *($p<.05$), **($p<.01$) and ***($p<.001$). # means no significance test was possible due to missing data on shared variable to be predicted.

Based on C. Camerer, "General Conditions for the Success of Bootstrapping Models," *Organizational Behavior and Human Performance*, 27 (1981): 411-422.

experts' judgments performed at least as well as the experts themselves and often better.[13]

What it all adds up to is that doing something systematic is almost surely better than purely intuitive prediction. Indeed, even giving equal weight to the most important predictors usually outperforms intuition.[14] These techniques get you the benefit of perfect consistency and limited distortion at the cost of an upfront investment in time and effort. Of course, for repeated decisions, this investment can be amortized.

Implementation—Just about anyone who knows a spreadsheet program can set up a table to calculate the simpler form of importance weighting. You will need to assign one of your statistics experts to create the more complicated regression model that determines implicit weights. But the technical details of implementing these methods are not as difficult as the organizational issues.

At Harris Investment Management, a unit of the Harris Trust and Savings Bank in Chicago, management was concerned that although the analysts and portfolio managers often had good ideas about investment strategies, they were at times distracted by recent market information or the current strategy's poor performance. So Harris decided to create a model that would combine its experts' insights (e.g., about the yield curve, economy, specific industry sectors) to guide its overall investment strategy. Use of the models ultimately improved the company's bottom-line results, but it was not easy.

First, analysts and portfolio experts had to be persuaded that their intuitive judgments might not be totally free of bias (a delicate matter). Second, these experts needed to accept the model that combined their insights as their friend, not their rival. Third, the experts wanted the power to override the model in case its use would be clearly inappropriate (as during the 1987 stock market crash).

You may encounter considerable reluctance and skepticism on the part of your experts. Try to persuade them by emphasizing the dangers of purely intuitive or rule-based judgments. Then present the model as no more than a computational encapsulation of their considerable expertise. Don't give the model too much of a separate identity, lest it be viewed as a competitor. Emphasize that the model cannot run without their inputs and may be overridden by them if circumstances warrant. Lastly, keep track of the model's performance, improve it, and slowly persuade your team that everyone is better off combining intuition and analysis rather than relying on one approach alone. Depending on the strength of the bootstrap effect, the role of your experts (and their attitude toward the model) may vary considerably. At Harris, several intuitive experts left the bank because they perceived their role as having been "diminished" by the new process.

Each organization must find its own optimal balance for layering intuitive

and analytic approaches. Sometimes a 50-50 combination of a bootstrap model and your top experts is best.[15] Harris uses a combination to select bonds. A bootstrap model takes into account factors such as industry sectors, maturation profile, and interest rates, whereas analysts and portfolio managers consider trading liquidity, special bond features, ethical investment constraints, and new information.

But even bootstrap models have their limitations. They don't consider how the factors are linked to ultimate goals and strategies, and they assume that an increase in a given factor adds value at a constant rate. Value analysis addresses these issues.

Value Analysis

> I cannot for want of sufficient premises, advise you what to determine, but if you please I will tell you how.
> —Benjamin Franklin[16]

When a decision is truly important and complex, it may pay to conduct a more comprehensive assessment. Value analysis refines importance weighting techniques by considering how factors affect broader objectives and how increases in the rating of a factor add value. The technical details can be quite complicated, and you will need to channel much of the detail work to experts in decision analysis, but the concepts are not difficult. If you know that such an option exists, you will be able to gather the necessary resources when it's time to make that big, important decision.

Key Objectives—Value analysis goes beyond lists of factors to uncover the true *values* of the decision maker. It does this by linking the factors to key objectives. For instance, in the admissions case, you would start by determining the broad characteristics you are looking for in your applicants, such as intellectual ability, professional commitment, and leadership potential. Then you would use scientific methods to determine how well GMAT scores, undergraduate major, and so forth actually predict those three criteria.

One of the present authors used value analysis to help a large oil and gas multinational make an important strategic investment decision. The company needed to build a $500 million pilot plant to test a process that could convert natural gas into middle distillate fuels. The key question was in which country to locate this first commercial plant. More than 10 countries were considered (from Germany to Malaysia to New Zealand), and they had dozens of advantages and disadvantages. In some countries, the investment climate was very attractive; in others labor costs were low or the supply of natural gas was abundant.

To make the decision, the executives identified four overriding objectives: financial attractiveness, the degree of uncertainty and risk, strategic fit,

and organizational desirability. They connected the features of specific countries with these objectives in a "goal hierarchy," as summarized in Figure 1. Senior executives determined the weights for the higher-level objectives, and technical experts estimated and weighted the lower-level ones. Because the analysis was complete, careful, and generally honest, the company reached agreement on the overall ranking among the ten possible countries and built the plant in the country that came out on top.

Nonlinear Values—Value analysis also addresses the fact that an increase in a given factor does not necessarily add value at a constant rate. In the business school admissions case, for example, twenty years of work experience is not necessarily twice as valuable as ten years of work experience. At some point, work experience contributes to overall performance in smaller and smaller amounts. In statistical terms, the value function is nonlinear.

Another example is GPA. You may believe that an increase in GPA from 3.6 to 3.8 is much more impressive than an increase from 3.0 to 3.2. Value analysis techniques can accommodate these changes in factor value.

Bring on the Experts—The technique that combines these refinements is called multi-attribute utility (MAU) analysis. This is where a trained decision analyst comes in. The process is quite costly in terms of time and effort, requiring one or more days with the analyst.

Again the difficult implementation problems are going to be organizational rather than technical. Many people will have trouble understanding how the method helps with decisions. Only highly numerate types will understand the theory and calculations. But a good analyst should be able to translate any technical aspect into lay terms. Once an MAU model is built, it can then handle most of the remaining informational and computational complexities. The final decision will be highly explicit in terms of the data employed, assumptions made, and value weights assigned. If done right, it should withstand a fierce public or private inquiry.

Is It Worth It?—Does all this refinement yield better decisions? Many analysts think so. MAU analysis has been applied effectively to decisions about locating airports and nuclear power plants, prioritizing fire department services, and setting corporate objectives.[17] One company used it to decide which of three prototype testers for very large-scale integrated circuits to bring to market.[18]

But more generally, there is no one answer. The issue boils down to a subjective tradeoff between effort and decision quality. Is one more day of analysis (and the associated costs) worth a 10% increase in the quality or defensibility of your next decision? If big bucks are at stake, it probably *is* worth it.

Figure 1. Value Analysis Using a Goal Hierarchy
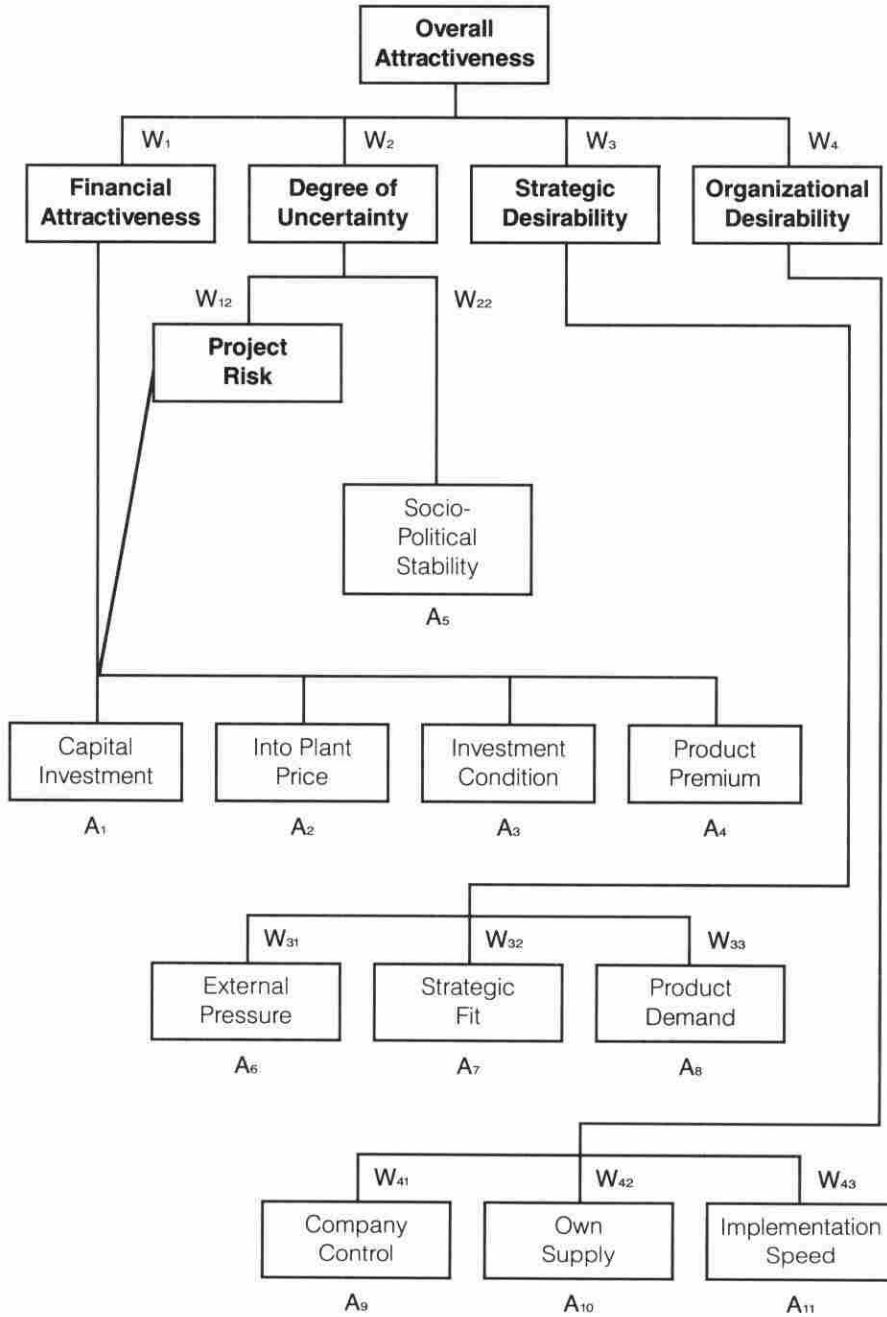
A Pyramid of Decision Approaches

The four methods we have described can be illustrated in a pyramid, as shown in Figure 2. The higher the method is on the pyramid, the more accurate, complex, and costly it tends to be. We use this shape to show that higher approaches are used less frequently than lower ones and for more important decisions.

To illustrate how the different methods can lead to different results, we applied them to the decision many MBA students must make: where to relocate upon graduation. First we asked our students to make a list of their top 10 U.S. cities for relocation, assuming no differences in job attractiveness and cost-of-living adjusted salary. They made intuitive decisions based on information from the *Almanac of Places Rated*, which scores 97 U.S. cities in terms of climate, housing cost, health care, environment, crime, transportation, education, recreation, the arts, and economics.[19]

We also gave each student a personal computer software package that was programmed with procedures for threshold rules, dictionary rules, and value analysis.[20] The students chose their own thresholds and tightened or loosened them until the program came up with 10 cities. For the dictionary rule, students ranked the attributes in order of importance and indicated for each how small a difference would constitute a tie. Whenever a tie occurred, the program would go to the next most important attribute.

The MAU procedure required students to assign each attribute a weight and determine how increases in a factor's rating added value. For example, an increase in the average temperature might be greatly valued up to, say, 80°F, and then flatten or decline in attractiveness. The computer program then performed the calculations and identified the top 10 cities.

Figure 2. A Pyramid of Approaches
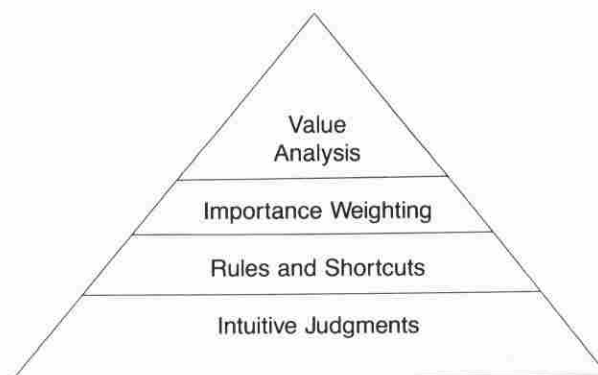
Each student thus generated four lists of 10 cities. In Table 3, we show one student's lists and the average percentage of agreement between lists for all the students. On average, the MAU method differed from the others about 50% of the time. What accounts for these differences? Simply the fact that the methods process and weight information differently. For instance, the student whose selections are listed in Table 3 may have romantic ideas about living in Honolulu (which appears on the intuitive list but nowhere else) that have nothing to do with his actual preferences for climate, cost of living, and so forth. The threshold rule is apparently knocking out San Francisco because the city doesn't pass the test on a few criteria (such as cost of living and safety), although the other three lists include it. And the dictionary rule includes Milwaukee; apparently that city scores high on a few important (such as being near one's family) criteria but not others, because it doesn't appear on the other lists.

What this experiment shows is that the *different* techniques do result in *different* decisions for the same problem; the selection of method matters. The key question is whether a particular decision requires a high level of accuracy. Will this MBA student be unhappy if he moves to Honolulu? Perhaps—when he finds out, say, how much it costs to buy a house there and how far it is from the mainland.[21]

Choosing an Appropriate Method

Each of the four techniques we have described has its advantages and disadvantages (see Table 4). Intuition is notoriously unreliable, but because it takes so little effort, it may be appropriate for some decisions. In an emergency, there may be no time for a comprehensive assessment. Or the decision may be inherently intuitive, such as artistic judgments.[22] But don't be too quick to use intuition even in these cases. Orley Ashenfelter, a Princeton University economist, developed a model to predict the quality of Bordeaux wines[23] Usually, wine experts carefully taste samples of a new year to predict its ultimate quality. To their surprise and dismay, a simple three-factor equation (winter rainfall, average summer temperature, and harvest rainfall) does just as well.

Rules take a little more effort, provide a bit more quality, and are moderately transparent (that is, they are easy to use and can be used to defend a decision, to some degree). Importance weighting takes a lot of effort the first time, but it can be used quickly thereafter. It yields high-quality decisions and very high clarity. Value analysis, while providing the highest-quality decisions, requires a maximum level of effort and is very difficult to explain and use.

To make these tradeoffs, you must consider the *importance* of the decision at hand. How high are the stakes? How frequently will the decision be made? Improving seemingly minor decisions that are made frequently, like

Table 3. City Selection via Four Different Methods

(a) One Student's Top Ten Selections by Method

Intuitive Judgment	Threshold Rule	Dictionary Rule	MAU Model
San Francisco	Boston	Seattle	Seattle
Honolulu	Buffalo	San Francisco	San Diego
Boston	Cincinnati	Cleveland	San Francisco
Chicago	Cleveland	Chicago	Cleveland
Nassau	Columbus	Milwaukee	Chicago
Sacramento	Pittsburgh	San Diego	Washington
Denver	Portland	Cincinnati	Portland
San Diego	San Diego	New York	Pittsburgh
New Brunswick	Seattle	Baltimore	Boston
New York	Washington	Dallas	Cincinnati

(b) Agreement Among Rules When Averaged Over All Subjects

	Threshold Rule	Dictionary Rule	MAU Model
a. Intuitive Judgment	37%	42%	51%
b. Threshold Rule	X	41%	50%
c. Dictionary Rule		X	55%
d. MAU Model			X

Table 4. Pros and Cons of Different Approaches to Pyramid

Method Used	Quality	Effort	Clarity
1. Intuition	Low	Low	Very Low
2. Rules	Moderate	Little	Moderate
3. Weighting	High	High/Low	Very High
4. Value Analysis	Very High	Very High	Often Low

loan or admission decisions, may do as much good over time as making one big decision well, such as choosing the next CEO.

You must also consider the *complexity* of the decision. Is the information involved so complex that the decision makers suffer from mental overload? Are the deeper or core values that underlie a decision difficult to articulate?

Other importance or complexity dimensions may influence a decision, such as time pressure, resource constraints, political ramifications, or issues of justification.[24] Generally, intuition makes sense for small decisions with limited information complexity (e.g., whether to attend a conference or

donate to a worthy cause). Also, it may be appropriate for cases where expertise has become truly "automated" in someone's mind through years of training and experience. For most other business decisions, we recommend increasing levels of sophistication as a function of decision importance and complexity.

Implementation

You have two tasks as you attempt to spread these techniques through your organization: making the technical tools available and encouraging people to use them.

In large organizations, enough in-house expertise may exist to examine all four approaches. Statisticians, operations research people, or market researchers should be able to devise tests to measure the accuracy and distortion in intuitive judgments. They can also perform the regressions and perhaps would be able to apply value analysis. For smaller firms, outside consultants are indicated, either from academia or quantitatively oriented management consulting firms. Few consultants, however, will be familiar with both the behavioral and quantitative literatures referred to in this article.

Of course, organizational resistance is likely to be a much bigger problem than developing the technology. Even after our MBA students had tested their intuition on the admissions cases (with dismal results), they still trusted intuition almost as much as they trusted MAU analysis in the city selection experiment.

Here are some practical suggestions for disseminating better decision-making techniques:

- Experiment with all the approaches we discussed; make people aware that they use methods at different levels of the pyramid and that each level has its place. During your next retreat, when some tough choices come up, split the group into three teams. Ask one team to tackle the issue intuitively, another team to develop a rule, and a third team to build an importance weighting model. Then see if they come back with different rankings, and, if so, ask them why.
- Identify where important decisions in your organization are currently being made on purely intuitively grounds. Check the track record of these intuitive experts and see if you can perform a test-retest study concerning their consistency. Are some of your experts perhaps as unreliable as the X-ray specialists or our MBA students? If so, consider bootstrapping.
- Perform a "rule audit." What rules are being used in your organization and with what results? Do the people who use them appreciate the rules' distortions? Ask for cases where the rule would be greatly off the mark.

Work with these decision makers to fine-tune rules or adjust them to changes in the general environment.

- Explore opportunities for modeling in such areas as personnel, new product design, compensation, sales prediction, budget estimation, selecting R&D projects, and so on. Encourage your experts to create models based on their own criteria and then challenge them to outperform their own models.

- Discuss explicitly with your staff the tradeoffs you are willing to make among consistency, distortion, speed, and clarity. Perfection has its price, and value analysis should be reserved only for the more complex and important decisions. Help decision makers consider those tradeoffs and lead them to appropriate method selections.

In the future, we expect that higher-effort, higher-quality approaches, especially bootstrapping models, will become more common in organizations. This is because: a growing body of behavioral research is documenting the hidden dangers of lower-quality techniques; technical advances in decision software and artificial intelligence are being made that facilitate the use of higher-level techniques; and competitive pressures to improve decision quality are increasing. Managers that disseminate these methods now will build quality decision making into the very fabric of their organization.

References

1. M.J. Prietula and H.A. Simon, "The Experts in Your Midst," *Harvard Business Review* (January/February 1989), pp. 120-124.
2. P.J. Hoffman, P. Slovic, and L.G. Rorer, "An Analysis-of-Variance Model for Assessment of Configural Cue Utilization in Clinical Judgment," *Psychological Bulletin*, 69 (1968): 338-349. To arrive at 23%, we took the reported mean infra-export correlation of .76 and calculated the chance of getting a reverse opinion between two cases from one time to the next. In general, a correlation of r translates into a $[.5 + \arcsin(r)/\pi]$ probability of a rank reversal of two cases the second time, assuming bivariate Normals [see M. Kendall, *Rank Correlation Methods* (London: Charles Griffen & Co., 1948)].
3. J.E. Russo and P.J.H. Schoemaker, "Managing Overconfidence," *Sloan Management Review*, 33/2 (Winter 1992): 7-17.
4. R.J. Ebert and T.E. Kruse, "Bootstrapping the Security Analyst," *Journal of Applied Psychology*, 63 (1978): 110-119.
5. Correlation is a statistical measure that ranges from $+1$ (perfect agreement) to 0 (no agreement) to -1 (complete reversal).
6. J.W. Payne, J.R. Bettman, and E.J. Johnson, "The Adaptive Decision Maker: Effort and Accuracy in Choice," in R.M. Hogarth, ed., *Insights in Decision Making* (Chicago, IL: The University of Chicago Press, 1990), pp. 129-153.
7. R.M. Dawes and B. Corrigan, "Linear Models in Decision Making," *Psychological Bulletin*, 81 (1974): 105.
8. See T.L. Saaty, *The Analytic Hierarchy Process* (New York, NY: McGraw-Hill, 1980).
9. J. Dowey, "Pinto Madness," *Mother Jones*, 18 (September/October 1977).

10. H. Wainer, "Estimating Coefficients in Linear Models: It Don't Make No Nevermind," *Psychological Bulletin*, 83 (1976): 213-217. Nonetheless, cases exist where the choice is sensitive to the exact weights. In our own research [P.J.H. Schoemaker and C.C. Waid, "An Experimental Comparison of Different Approaches to Determining Weights in Additive Utility Models," *Management Science*, 28/2 (1982): 182-196], we found that the different methods often do yield different weights but that the choices or predictions resulting from those weights are usually not significantly different.

11. Ebert and Kruse, op. cit.

12. R.M. Dawes, D. Faust, and P. Meehl, "Clinical versus Actuarial Judgment," *Science*, 243 (1989): 1668-1673.

13. Of the 15 studies listed in the table, seven showed a statistically significant ($p<.05$) difference between intuitive performance and the model. Since both correlations involve the same target variable, a z-test for correlations from dependent samples was used. The fact that 14 of the 15 studies show a positive bootstrapping effect (and none are negative), is itself highly significant under an exact binomial test.

14. Dawes and Corrigan, op. cit.; H.J. Einhorn and R.M. Hogarth, "Unit Weighting Schemes for Decision Making," *Organizational Behavior and Human Performance*, 13 (1975): 171-192; Wainer, op. cit.

15. R. Blattberg and S. Hoch, "Database Models and Managerial Intuition: 50% Model + 50% Manager," *Management Science*, 36/8 (1990): 887-899.

16. B. Franklin, "Letter to Joseph Priestley," 1772, reprinted in *The Benjamin Franklin Sampler* (New York, NH: Fawcett, 1956).

17. R.L. Keeney and H. Raiffa, *Decisions with Multiple Objectives* (New York, NY: Wiley, 1976); R.L. Keeney, *Value Focused Thinking* (Boston, MA: Harvard University Press, 1992).

18. R.L. Keeney and G.L. Lilien, "New Industrial Product Design and Evaluation Using Multiattribute Value Analysis," *Journal of Product Innovation Management*, 4, (1987): 185-198.

19. R. Boyer and D. Savageau, *Places Rated Almanac* (New York, NY: Prentice-Hall, 1984).

20. J. Hershey, H. Kunreuther, and S. Schocken, "Integrating Prescriptive and Descriptive Analyses Using A Computer-Based Decision Support System," Working Paper, Department of Decision Sciences, The Wharton School, University of Pennsylvania, April 1984.

21. Of course, we don't know for sure that the MAU method indeed identified the best options for each student. It would require many hours with each person to model his or her true preference in fine detail. However, the MAU method used in this experiment comes closest to such a full-blown value analysis.

22. K.R. Hammond, R.M. Hamm, J. Grassia, and T. Pearson, "Direct Comparison of Efficacy of Intuitive and Analytical Cognition in Expert Judgment," *IEEE Transactions on Systems, Man, and Cybernetics*, 5 (September/October 1987), SMC-17.

23. P. Passell, "Wine Equation Puts Some Noses Out of Joint," *New York Times*, March 4, 1990, p. 1; O. Ashenfelter, D. Ashmore, and Robert Lalonde "Wine Vintage Quality and the Weather: Bordeaux," Working Paper, Department of Economics, Princeton University, February 1993.

24. P. Kleindorfer, H. Kunreuther, and P. Schoemaker, *Decision Sciences: An Integrative Perspective* (New York, NY: Cambridge Press, 1993).